Examining Child Care Subsidy Receipt: An Analysis of Matched NSECE and Illinois Administrative Data

OPRE

NSECE ANALYTICAL REPORT

OPRE Report #2016-12 | FEBRUARY 2016

National Survey of **Early Care & Education**

Disclaimer

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

Examining Child Care Subsidy Receipt: An Analysis of Matched NSECE and Illinois Administrative Data

NSECE ANALYTICAL REPORT
OPRE REPORT 2016-12 | FEBRUARY 2016

Submitted to:

Ivelisse Martinez-Beck, Ph.D., Project Officer
Ann Rivera, Ph.D., Research Analyst

Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Numbers: HHSP233201500048I, HHHSP23320095647WC

Project Director:
A. Rupa Datta
NORC at the University of Chicago
55 E. Monroe Street
Chicago, Illinois 60603

National Survey of Early Care and Education Project Team (2016). Examining Child Care Subsidy Receipt: An Analysis of Matched NSECE and Illinois Administrative Data. OPRE Report #2016-12, Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

NSECE Research Team

This is a collaborative product of the NSECE research team. Key contributors for this report were:

Wladimir Zanoni
Robert Goerge, Principal Investigator
A. Rupa Datta, Project Director
Lisa Gennetian
Ann Witte
Richard M Brandon
Martha Zaslow
Carolina Milesi
Lina Guzman

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre/index.html.

OPRE

NORC at the UNIVERSITY of CHICAGO

OVERVIEW

Since the Child Care Development Fund (CCDF) program was initiated in 1996, one key challenge faced by researchers has been data limitations for accurately measuring the enrollment of eligible households in this program while also having additional information about households for informed analyses. Combining survey data with administrative records can open new avenues for CCDF subsidy research. This report discusses new challenges and opportunities that arise as researchers combine survey data with CCDF administrative records to conduct subsidy research.

The National Survey of Early Care and Education (NSECE) team undertook an innovative approach to calculate CCDF program participation. Using probabilistic record linkage methods, the household records from the NSECE were matched to CCDF administrative data from the State of Illinois to form a combined database of survey and administrative data. That combined database allowed creation of CCDF program participation variables from NSECE households' over-time records in the childcare subsidies program. The unified database created from this exercise resembles one from a cross-sectional survey that, by asking retrospective questions, identifies households' recent participation in social programs (CCDF, in our case). But the unified database has the advantage of a more accurate participation variable from administrative data than would have been obtained from a survey self-report.

We highlight two methodological lessons for linking survey data with administrative data to study the CCDF program. First, the fact that researchers will almost always have consent from partial (rather than full) samples of households for linkage gives rise to two challenging statistical problems: potential bias and likely undercount. Second, the seemingly straightforward process of pairing the time of the HH-NSECE interview with that of subsidy receipt in administrative records demands careful conceptualization of what the NSECE variables express when they are measured after and/or before CCDF subsidy receipt.

The report demonstrates that linkage of files opens up substantial opportunities for new research. Through comparisons of CCDF receipt indicators referring to two different time periods, the report also suggests implications for subsidy research traditionally conducted using only administrative records.

BACKGROUND AND MOTIVATION

Understanding what factors affect participation in the Child Care and Development Fund (CCDF) program, and how it affects parental and children's outcomes, requires high-fidelity measures of participation of child care subsidy receipt. Since CCDF is one among various other funding streams for early care and education programs available to low income families, it is often difficult for survey respondents (households and providers alike) to correctly identify their participation in this program. Household surveys that include questions about CCDF subsidy receipt impose a high cognitive burden on respondents, and, for this reason, the accuracy of CCDF participation variables are often questioned.[1] In studying the characteristics of the eligible population, the determinants of the CCDF program participation, and the effects of this program, the early care and education field faces the challenge of finding data that accurately measures household participation in this subsidy.

The NSECE team undertook an innovative approach to calculate CCDF program participation. Using probabilistic record linkage methods, NSECE household records were matched to CCDF administrative data from the State of Illinois to form a combined database of survey and administrative data. That combined database recovers longitudinal histories of NSECE households in the childcare subsidies program, from which CCDF program participation variables were created. The unified database created from this exercise resembles one from a cross-sectional survey that, by asking retrospective questions, identifies households' participation in social programs (CCDF, in our case).

In this report, we document some methodological challenges encountered, and lessons learned in linking NSECE households across the two datasets, and in defining CCDF subsidy participation. In order to illustrate how to employ the unified database in applied research, we present an exploratory study of the associations between selected household characteristics and CCDF program participation.

The insights provided in this report are possible because the NSECE secured respondents' consent to match to administrative records. This consent allows researchers to combine high-fidelity measures of childcare subsidy utilization from administrative records with exceptionally detailed information on household characteristics from the NSECE survey. This report provides a road map for other researchers who may wish to employ similar methods using the NSECE or with other pairings of survey with administrative data.

This report is divided in three sections. In the first section we briefly describe the data. The second section describes methodological aspects related to the exercises of linking the two datasets (NSECE Household Survey and the Illinois' CCDF administrative data) and of constructing a CCDF subsidy receipt indicator. The last section describes how we employ the linked data to conduct a basic descriptive analysis of the associations between selected household characteristics and CCDF utilization.

[1] Cognitive interviews conducted by the NSECE suggest that respondents cannot accurately identify their CCDF enrollment. The CCDF program works as a funding mechanism and can exhibit a high degree of collaboration with multiple other ECE programs (such as state-funded pre-K and Head Start/Early Head Start). In addition, States have substantial leverage to define the program's features within the eligibility rules set by the Federal government. While parents can identify where their children attend ECE, the funding source(s) can be rather unknown to them.

DATA

CCDF Data from the State of Illinois

The Illinois' Child Care Tracking System (CCTS) is a relational longitudinal database of administrative records that tracks CCDF subsidies payments in Illinois since the creation of the program in 1997. Every month CCTS data identify all children who have received CCDF-related payments, and links that payment information to their child care providers (data on providers includes their types of care and addresses). The database includes basic demographic characteristics, family composition data and addresses of children and households participating in the CCDF program. Details of the CCDF program in Illinois (the Child Care Assistance Program—CCAP) are provided in Section A of the Appendix.

The NSECE Household Survey (HH-NSECE) Database

Collected from January to June 2012—a period of relative stability post the Great Recession, data from the NSECE include a household survey with extensive information on socio-demographic characteristics, as well as detailed records of parental work and children's exposure to early care and education. In this report we analyze a subsample of the HH-NSECE survey that corresponds to respondents in the State of Illinois. With these data linked to the CCTS database described above, we investigate correlates of childcare subsidy utilization.[2]

Record Linkage across Survey and Administrative Data

Using probabilistic record linkage methods, children's records with identifying information from the HH-NSECE survey (first name, last name, date of birth and street address) were linked to records in the Illinois Childcare Tracking System (CCTS). The basic idea was to probabilistically identify the same entities (children and households) across the two databases on the basis of similarities of their identifying information. The procedure built a probability model to pair records of children's first names, last names, dates of birth and residential address across the two databases (NSECE and CCTS) on the sample of NSECE children (and consequently households) for whom consent to be matched to administrative records was granted by the survey respondent.

The probabilistic record linkage methods that we employed allowed us to assign match-weights to each matching variable (i.e., first names, last names, dates of births, and residential addresses). Matching variables that are believed to be more accurate and/or important were assigned higher match-weights: household respondents' first names, last names, and dates of birth received the highest match-weights; while residential addresses—a time-variant field which we have for only one point in time-- received lower match-weights. Balancing match-weights in this way ensures that we optimally use all the relevant information to uniquely identify individuals across the two databases for linkage. Match-weights were used solely in the linking process and not for subsequent analyses using the linked data.

[2] Please see the NSECE Summary Data Collection and Sampling Methodology Report for more information about the study. National Survey of Early Care and Education Project Team (2013). *National Survey of Early Care and Education: Summary Data Collection and Sampling Methodology*. OPRE Report #2013-46, Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

METHODOLOGICAL CONSIDERATIONS

In this section we describe two methodological aspects of the present research: dealing with self-selection in survey consent, and defining child care subsidy participation. The NSECE survey asked respondents for their consent to use identifiable information of HH members for matching to administrative records. Determining whether the sample of respondents is a random one, and dealing with a self-selected sample were important tasks in this analysis. While defining childcare subsidy participation may appear straightforward, linking survey data to administrative data involves challenges that result from the need to pair the time of the interview with that of subsidy receipt. The objective of this section is to provide a methodological guide for future research that aims at linking survey and administrative data to study child care subsidies.

Survey Consent and Childcare Subsidy Receipt

The NSECE Household Survey questionnaire asked for parental consent to use identifying information of household members so that researchers could match their data to governmental administrative records. As Exhibit 1 shows, not all respondents gave this consent to the NSECE interviewers. In Illinois, 570 out of 760 survey respondents (about 75 percent of households) authorized the research program to use their personal identifying information. Nationally, NSECE respondents granted consent-to-match for approximately 73 percent of children.

Exhibit 1. Counts of NSECE Households in Illinois, by Consent to Link to Administrative Data

	Household Counts	
	All NSECE HHs in Illinois	NSECE IL Households Giving Consent
Unweighted	760	570
Weighted	1,190,000	898,000

Note: Unweighted counts are rounded to the nearest ten (10) for disclosure protection purposes.

Important challenges for this research arise from the fact that we can only make inferences from the sampled NSECE Illinois population who granted consent to be matched to administrative records. It is important to determine whether the subsample of NSECE respondents who gave consent exhibits systematic differences in characteristics that distinguish them from the subsample of households who did not give matching consent. A direct consequence of "non-random sorting" of households across the consent dimension is that estimates computed solely on data from households who gave consent would not be representative of the full sampled population (in our case, the Illinois population of households with children under age 13 years). Making inferences from a self-selected sample of households who gave consent would be especially problematic (i.e., prone to bias) if variables showing systematic differences across the consent dimension matter for predicting CCDF subsidy receipt.[3]

Another consequence of having consent to match for only part of the sample is that the working sample of households becomes a truncated one. Given that the underlying number of households that consented is smaller than the sampled number of households, sampling

[3] If the households that gave consent do not differ systematically from those who did give consent, then analyzing exclusively the data from households that consented would reduce sample size impacting count estimates, yet will have not bias implications for estimates of means and other distribution moments. As we will see below, this seems to be the case for the data in this analysis.

weights need to be adjusted. Adjusting the sampling weights is needed so that the weighted number of observations in the subsample of HHs who give consent describes the full target Illinois population (and not a subset of it). Without adjustments to the weighting scheme, estimated counts based on the truncated sample will not capture the full underlying Illinois population, which is our population of interest. In other words, even when estimates of means, medians and other moments are likely to be representative of both populations, if there are systematic differences between households that consent and those that do not, the counts of households falling within specific categories will not be accurate.

In order to deal with the partial consent-to-match problems outlined above, we first evaluated whether households systematically differed in key observable characteristics across the consent dimension.[4] To study this pattern, we employed a regression framework and tested for both the individual and the joint statistical significance of selected household level variables in predicting survey consent. The survey consent measure that we used was an indicator variable that turns one if the NSECE respondent gave consent for at least one child, and zero otherwise. Using multivariate linear probability, Probit and Logit regression, we modeled the consent decision as a dependent variable, explained by categorical variables measuring presence of children younger than five years old in the household, income to poverty ratios, respondents' educational attainment and race/ethnicity.

In dealing with the income data, we found that the HH-NSECE poverty variable in our analysis had some missing values. The poverty variable from the survey measures the 2011 household income to federal poverty line ratio (or income to poverty bracket). Out of the 760 households in the Illinois sample, 30 of them (approximately 4%) have missing values in this variable. The missing data appear unexplained by race/ethnicity, presence of children less than 5 years old, or the respondent's education level (as we determined it using multivariate regression-based analysis). Missing data in the poverty variable are therefore assumed at random (see Section B of the Appendix for details).

[4] We make the assumption that consent is based solely on observable attributes (and not unobservables) that the researchers can measure accurately.

Exhibit 2. Modeling the Household Consent Decision

Variables	(1) OLS	(2) LOGIT	(3) PROBIT
One+ children aged 0 to < 60 months	0.0273	0.1557	0.0896
	(0.0319)	(0.1794)	(0.1050)
8th grade or less[1]	-0.1007	-11.9837	-3.3590
	(0.4293)	(1186.6816)	(154.7285)
9th- 12th grade[1]	-0.2562	-13.0835	-3.9674
	(0.4269)	(1186.6816)	(154.7283)
HS or GED[1]	-0.2083	-12.8262	-3.8160
	(0.4266)	(1186.6816)	(154.7283)
Some college/no degree[1]	-0.2030	-12.7984	-3.8003
	(0.4275)	(1186.6816)	(154.7283)
Associate degree[1]	-0.2256	-12.9235	-3.8762
	(0.4294)	(1186.6816)	(154.7284)
Bachelor degree[1]	-0.2447	-13.0219	-3.9334
	(0.4291)	(1186.6816)	(154.7284)
Graduate degree[1]	-0.1961	-12.7628	-3.7783
	(0.4325)	(1186.6816)	(154.7285)
<100% poverty[2]	0.0037	0.0097	0.0087
	(0.0545)	(0.2996)	(0.1765)
>=100 to <200% poverty[2]	0.0341	0.1917	0.1151
	(0.0526)	(0.2920)	(0.1701)
>=200 to <300% poverty[2]	0.0056	0.0265	0.0218
	(0.0604)	(0.3281)	(0.1938)
White[3]	-0.0363	-0.2008	-0.1180
	(0.0518)	(0.2895)	(0.1691)
Black[3]	-0.0317	-0.1715	-0.1012
	(0.0550)	(0.3070)	(0.1800)
Hispanic[3]	-0.0051	-0.0256	-0.0177
	(0.0532)	(0.3007)	(0.1749)
Observations	730	730	730
Log likelihood	-402.5106	-391.0550	-391.0695
F –statistic	0.6556	-	-
χ2 –statistic	-	10.2743	10.2454
P value joint significance	0.8183	0.7419	0.7440

Notes: The values in the upper panel of the table are coefficient estimates from regression models. Standard errors in parentheses. The dependent variable is an indicator variable turning one if the household gave consent to be matched in administrative records. Asterisks denote: * p<0.05 ** p<0.01 *** p<0.001". [1] The reference group for the educational attainment categories is “No education.” [2] The reference group for the 2011 income to poverty ratio categories is “>=300% poverty”. [3] The reference group for the race ethnicity categories is “Other non-Hispanic race/ethnicity”.

Exhibit 2 presents parameter coefficients from OLS and Logit and Probit binary regression models of NSECE survey respondent consent. In the table, we can observe that respondents' consent is not predicted accurately by any of the variables included in the analysis (neither individually—based on the T-statistic- nor jointly—based on either the F or the Chi Squared statistics). These findings suggest that the data generating process underlying respondents' decisions to provide consent to be matched to administrative records follows a random pattern (random on observables). Consequently, in light of the lack of associations in the model between consent and the selected predictors, it is not necessary to develop a strategy to deal with biases in means arising from non-random consent-to-match.[5]

Even though there seemed to be no reason to be concerned about biases in means and other moments from analyzing a self-selected population, it was still important to deal with the fact that the underlying population associated with the households that gave consent in Illinois is smaller than the full sampled population. Consequently, the NSECE sampling weights still needed to be adjusted to reflect the full target Illinois population, and not to represent a random subset of it (that corresponding to consenting households). In order to do that we adjusted the NSECE Household Survey household sampling weights by the inverse of the probability of consent, which in this case, is a constant value. Section C of the Appendix gives details on how we implemented the inverse probability weighting scheme.

Exhibit 3 shows the estimated population counts of Illinois households before and after adjusting the sampling weights. The columns to the left reproduce the unweighted number of observations from the whole Illinois sample and the columns to the right show weighted counts for the whole population, the consenting sample with original sampling weights truncated and consenting sample with adjusted sampling weights.

Exhibit 3. Sample Sizes and Weighted Populations before and after Adjustment in Sampling Weights

	Unweighted counts		Weighted counts		
HOUSEHOLD CHARACTERISTIC	All HHs	Gave Consent	All HHs	Gave Consent (No Adjustment)[1]	Gave Consent (Weights Adjusted)[2]
NSECE Illinois	760	570	1,190,000	898,000	1,190,000

Notes: The numbers in the table represent population counts (weighted and unweighted).
[1] No adjustment indicates that the survey sampling weights have not been redefined to account for the fact that the sample of households who consent is smaller than the sampled population of Illinois households.
[2] "Weights adjusted" indicates that the sampling weights have been recalibrated to reflect the fact that the underlying population in the consented sample should be the target Illinois population of households with children under age 13 years.

The numbers in Exhibit 3 evidence the importance of adjusting the sampling weights to deal with the fact that the analytic sample (that of households that gave consent) is smaller than that of the Illinois population. The last two columns reflect a nearly 25 percent difference in weighted counts that would have underestimated the population size. Adjusting the survey weights in the

[5] If the sample of consenting households demonstrated bias on observable characteristics, we would have developed an inverse probability weights (IPW) approach. The intuition behind the IPW strategy is simple: assign larger weights to consenting families that are more similar in probability-of-consent to households that did not give consent. This may be necessary for other U.S. states or analytic populations of interest. Section C of the Appendix discusses IPW further.

way described is relevant to this analysis, given that one of our objectives is to determine the population of households participating and not in the CCDF program by several attributes.

Constructing a CCDF Subsidy Receipt Indicator

Conceptual Issues in Defining the Timing of CCDF Subsidy Receipt. The nature of the combined data described in the previous sections (the cross-sectional NSECE linked to the longitudinal CCDF administrative database) allows us to build measures of subsidy receipt to answer different research questions. Each research question, entails assumptions about how the timing or chronology of the HH-NSECE survey should be paired to the time in which survey respondents appear in the CCDF administrative records to create the CCDF participation measure.

For example, to study what factors predict CCDF participation, subsidy receipt should be computed as occurring *after* the NSECE interview. Constructed that way, researchers could claim that the NSECE survey measures temporally predetermine CCDF program participation, and, therefore, the NSECE measures could be claimed as potential predictors of subsidy use. Alternatively, to study associations between CCDF subsidy receipt and child, parental, and/or household outcomes in the HH-NSECE, CCDF receipt measures should be constructed *so* that their timing *precedes* the outcomes observed in the NSECE. If subsidy receipt precedes outcomes (such as household income or parental employment), there is a clear temporal ordering between the two.[6]

For this report, we chose to define a time dimension for program participation that was consistent with hypothetical survey questions of the type: “Have you used childcare subsidies during the last X months/weeks/days?” Thus, CCDF receipt precedes the measures observed in the NSECE data.

Notice that, from a conceptual standpoint, another approach to constructing CCDF subsidy participation variables is to have the survey question “mimic” administrative records. For instance, we could have constructed a CCDF participation measure consistent with a set of survey questions of the type: “Do you currently receive CCDF?” and, conditional on participation, to ask a follow up question: “For how long have you been receiving it?” We could have also followed households’ participation in childcare subsidies from the time of the NSECE interview prospectively, thus mimicking a longitudinal survey that samples households and inquires about CCDF participation periodically over time (such as in the Panel Study of Income Dynamics).

Empirical Issues in Implementing a Conceptual Definition of Child Care Subsidy Receipt. Two empirical considerations were relevant to pairing the NSECE interview date with administrative data to build our program participation indicator. Firstly, it was important to account for potential delays and/or inaccuracies in recording of payments in the CCDF administrative database. If CCDF payments are reported with delay, households may have received subsidies at the time of the interview, but their records dates may have lagged, inducing a “false negative” for participation in the program. Secondly, we needed to choose the retrospective time period prior to the interview during which CCDF receipt would be measured.

[6] However, this is still an analysis of associations between variables. Attributing causation to ‘CCDF effects’ would require other analytic techniques (such as instrumental variables) that remove endogeneity.

With this in mind, we used households' CCDF payment records to create a series of CCDF subsidy receipt indicators for different time intervals before and after the NSECE interview date.

Exhibit 4 shows unweighted and weighted counts of children and households participating in the CCDF program for combinations of days before and after the NSECE interview date. For instance, the cell in the first row and last column in Exhibit 4 shows how many households were found in administrative records considering a maximum of 30 days after the interview date for administrative delays, and given that the question of interest is whether the household received any CCDF payments within the entire year preceding the NSECE interview.

Exhibit 4. Defining Childcare Subsidies Take-up: Bandwidth Analysis, Counts of Children and Families Using CCDF

Unweighted number of children

		Days before survey date											
		30	60	90	120	150	180	210	240	270	300	330	360
	30	80	80	80	90	90	90	90	100	100	100	110	110
Days	60	80	90	90	90	100	100	100	110	110	110	110	110
after	90	90	90	90	90	100	100	100	110	110	110	110	110
survey	120	90	90	90	90	100	100	100	110	110	110	110	110
date	150	90	90	90	100	100	100	110	110	110	110	120	120
	180	100	100	100	100	110	110	110	120	120	120	120	120

Weighted number of children

		Days before survey date											
		30	60	90	120	150	180	210	240	270	300	330	360
	30	108100	119400	119400	130000	133100	133300	137700	141900	143400	143400	145300	147100
Days	60	113500	124300	124300	134900	138100	138200	142700	146900	148300	148300	150200	152000
after	90	113800	124700	124700	135300	138400	138600	143000	147200	148700	148700	150600	152400
survey	120	116700	124900	124900	135400	138600	138600	143000	147200	148700	148700	150600	152400
date	150	127100	135300	135300	145800	149000	149000	153400	157600	158100	158100	160000	161800
	180	135600	143700	143700	154300	156300	156300	160700	164900	165300	165300	167300	169100

Unweighted number of households

		Days before survey date											
		30	60	90	120	150	180	210	240	270	300	330	360
	30	50	50	50	50	60	60	60	60	60	60	60	70
Days	60	50	50	50	50	60	60	60	70	70	70	70	70
after	90	50	50	50	60	60	60	60	70	70	70	70	70
survey	120	50	60	60	60	60	60	60	70	70	70	70	70
date	150	60	60	60	60	60	60	70	70	70	70	70	70
	180	60	60	60	60	70	70	70	70	70	70	70	70

Weighted number of households

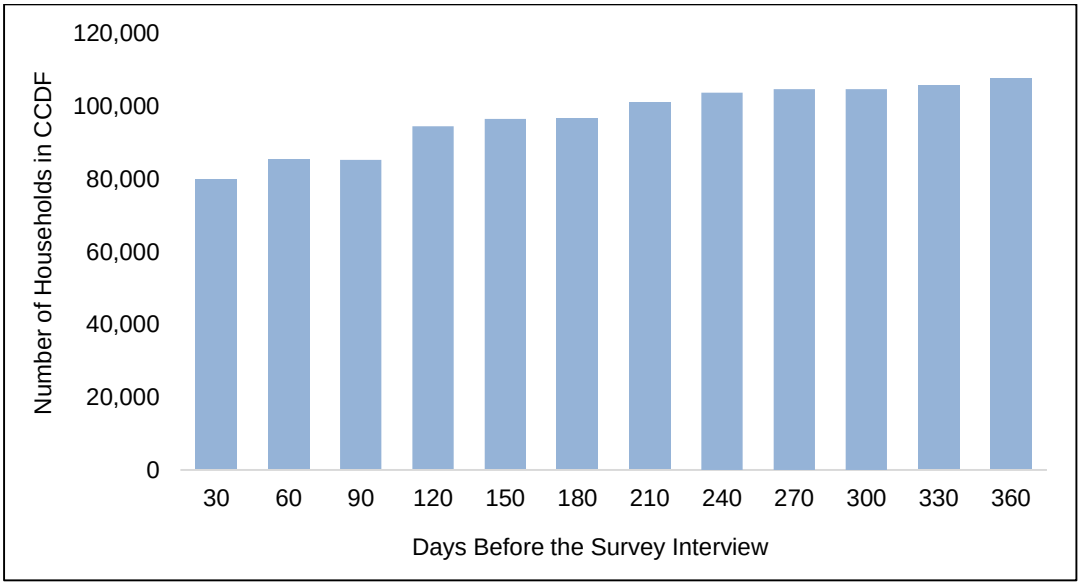
		Days before survey date											
		30	60	90	120	150	180	210	240	270	300	330	360
	30	76800	82300	82300	91500	93400	93600	98100	100700	101700	101700	102700	104500
Days	60	79500	84900	84900	94200	96100	96200	100700	103300	104300	104300	105300	107100
after	90	79800	85300	85300	94500	96500	96600	101100	103700	104700	104700	105700	107500
survey	120	81300	85400	85400	94700	96600	96600	101100	103700	104700	104700	105700	107500
date	150	91800	95800	95800	105100	107000	107000	111500	114100	114100	114100	115000	116800
	180	95700	99700	99700	109000	109800	109800	114200	116800	116800	116800	117800	119600

Notes: Unweighted numbers of children are rounded to the nearest ten (10). Weighted counts use adjusted weights to correct for (assumed) random consent-to-match (see previous section). Survey design effects are accounted.
(1) Days before the NSECE interview date vary according to how many days before the interview CCDF participation needs to be identified.
(2) Days after the NSECE interview date account for reporting delays in the Illinois' CCDF administrative data.

The bottom panel in Exhibit 4 shows that the numbers of households receiving CCDF subsidies increases as we increase the time intervals from which households are retrospectively retrieved in the CCDF administrative records. Exhibit 5 graphically depicts weighted counts of households receiving subsidies no more than 90 days after the NSECE interview date (i.e., allowing three months of delay in administrative data reports). Notice that the counts of CCDF participants increase at a decreasing rate as the retrospective time interval grows, with very little marginal

increase in participation after 240 days have been accounted for. Anchoring at 90 days after subsidy receipt, we observe an estimate of 79,800 families who receive CCDF within one month before the NSECE interview, and 107,500 families who participate no more than 360 days before the NSECE interview was conducted.

Exhibit 5. Households Receiving CCDF by Number of Days before the NSECE Interview



Notes: Weighted counts of households receiving CCDF subsidies include any household with payment records in the CCDF administrative database up to 90 days after the NSECE interview date. Data are taken from Exhibit 4 bottom panel-- highlighted row.

As we increase the posterior time interval (the number of days after the NSECE interview), we increase the risk of including people who start using the CCDF program after the date of the NSECE interview (i.e., the risk of false positives increases). Alternatively, by restricting the number of days accounted by the posterior time interval, we increase the probability of wrongly coding a household as not receiving CCDF when it is, in fact, participating in the program (i.e., the risk of false negatives increases). Ideally, the posterior time interval would be based on typical payment delays. Because we do not have this information, we used our judgement to settle on a 90-day posterior bandwidth for the subsequent analysis.

One final empirical issue is that our measure of CCDF participation is not entirely error-free; there are at least two remaining potential sources of measurement error. On the one hand, as already mentioned above, reporting errors in the administrative data might lead to error in the CCDF receipt indicator. Secondly, our procedure infers absence from the administrative data to mean “no receipt of CCDF.” While this is mostly true, it is important to note the possibility of false negatives due to poor matches across datasets. The extent and empirical implications of those two sources of measurement error are left for further analysis.

EXPLORATORY ANALYSIS OF LINKED DATA: ASSOCIATIONS BETWEEN HOUSEHOLD CHARACTERISTICS AND CCDF SUBSIDY RECEIPT

In this section we compare attributes of HH-NSECE households according to whether or not at least one of the children in those household receives CCDF subsidies. We evaluate the mean differences in selected variables from the HH-NSECE survey by CCDF receipt. The analysis was conducted for the whole analytic sample (all NSECE Illinois households who gave consent). In what follows, we employed two different pre-interview time intervals to define CCDF receipt: 30 days and 360 days before the HH-NSECE survey date. For either pre-interview time interval, we set a post-interview time interval of 90 days after the NSECE interview date to count households as receiving CCDF when their payment records have been delayed.

Differences in the proportion of households with specific attributes by CCDF receipt

Results shown in Exhibit 6 compare weighted proportions of household characteristics according to whether or not the households receive CCDF subsidies, and for two subsidy receipt variables that differ in their time dimension: CCDF receipt either 30 or 360 days before the survey interview.[7] In the exhibit we compare households in and out of the CCDF program across brackets of household income-to-poverty ratio, and by the educational attainment of the HH-NSECE respondent; the race and ethnicity of the survey respondent, and also the presences of at least one child younger than 60 months old, and at least one non-parent adult in the household; and the community poverty density and the degree of urbanicity of the neighborhoods where the households were located. Analysis of differences in means of the selected categories of variables by CCDF receipt status was conducted using pairwise t-tests.

[7] Weights are the adjusted sampling weights as described in the previous section; we also account for the NSECE sampling stratification and design effects.

Exhibit 6. Weighted Proportion of NSECE Illinois' Households by CCDF Participation

	CCDF up to 30 Days before HH-NSECE Date		CCDF up to 360 Days before HH-NSECE Date	
	(1)	(2)	(3)	(4)
	In CCDF	Not in CCDF	In CCDF	Not in CCDF
2011 income to poverty ratio in categories				
<100% poverty	0.505	0.248***	0.444	0.247**
100% to <200% poverty	0.369	0.234	0.445	0.222**
200% to <300% poverty	0.006	0.188***	0.022	0.191***
>=300% poverty	0.120	0.331*	0.089	0.339**
Educational attainment				
8th grade or less	0.017	0.053*	0.012	0.055**
9th- 12th grade	0.067	0.067	0.064	0.067
HS or GED	0.080	0.247***	0.132	0.246**
Some college/no degree	0.688	0.190***	0.564	0.190***
Associate degree	0.111	0.111	0.120	0.110
Bachelor degree+	0.037	0.225***	0.106	0.222**
Race and ethnicity				
White	0.327	0.552	0.243	0.566***
Black	0.449	0.118**	0.452	0.109***
Hispanics	0.161	0.175	0.170	0.174
Household composition				
One+ children aged 0 to < 60 months	0.644	0.559	0.547	0.566
One+ non-parent adult	0.406	0.379	0.431	0.375
Community poverty density in categories				
Low	0.633	0.760*	0.634	0.763*
Moderate	0.089	0.074	0.095	0.073
High	0.278	0.166	0.270	0.164
Urbanicity level in categories				
High density urban	0.871	0.928	0.902	0.926
Moderate density urban [a]	0.129	0.072	0.098	0.074

Notes: CCDF receipt defined when CCDF payment records are found for at least one child for whom consent was given. Subsidy receipt is calculated with reference to either 30 or 360 days before the survey interview as in the labels of the top row.*** Indicates that the mean difference of variable in the left row, for the sample indicated in the top row, by CCDF recipient status, is statistically significant at 99%.** Statistical significance at 95%. ** Statistical significance at 90%.

What Variables Appear Strongly Associated with CCDF Receipt?[8]

In Exhibit 6 we observe that, irrespective of the time interval used to define CCDF participation, income to poverty ratios appear to be associated with CCDF receipt. The comparison reveals that the income levels of subsidy recipients are lower than the income of those households who do not receive subsides. This result was expected, as we know that the CCDF program has strict income eligibility rules and targets low income families. The NSECE income-to-poverty ratio is based on the household's 2011 annual income. The NSECE-HH interviews took place between January and May 2012. Thus, the relationship of the poverty ratio to the two different CCDF receipt variables differs quite a bit. In the case of CCDF receipt within the 30 days prior to the NSECE interview, the 2011 income is likely to have preceded the entire 30 day-period prior to the measured CCDF receipt (although there may have been additional, earlier CCDF receipt during the 2011 calendar year).

The poverty ratio can be characterized as a predictor of subsidy receipt within the 30 days prior to the NSECE interview. However, in the 360-day case, the period for which CCDF receipt is measured (12 months beginning January through May, 2011) would overlap chronologically at least seven months, and possibly 12 months, with the calendar year 2011—which is the period for which income-to-poverty was measured. Consequently, the association between income and subsidy receipt in the 360 days case cannot be characterized as income predetermining CCDF receipt.[9]

A notable proportion of households in Exhibit 6 received CCDF subsidies either 30 or 360 days prior to the NSECE interview but had 2011 income to poverty ratios above 300 percent. The apparent discrepancy may be an indication of potential volatility in households' poverty levels, job loss, changes in family structure such as divorce or the birth of a new baby, or other factors that can cause a higher-income household in one year to become CCDF-eligible several months later. Other explanations may include eligibility for CCDF among higher-income households due to special needs of their children, or because (sub)family units measured for CCDF eligibility may differ from the household unit measured in the NSECE. Measurement error and other issues may also explain some of these discrepancies.

The numbers in the table indicate that educational attainment appears associated with CCDF receipt. Among students who were either current or recent graduates at the time of the NSECE interview, the measured CCDF receipt may have preceded the educational attainment we observe in the survey data. For most others, individuals' educational attainment will not have changed during or since the CCDF receipt recorded from the administrative data.

Exhibit 6 shows that when CCDF receipt is measured up to 30 days before the survey date, the proportion of CCDF participants who have up to HS or GED diploma (0.08) is lower than among non-participants (0.25). More importantly, perhaps, is the fact that the education levels of CCDF

[8] The unconditional comparison of outcomes (such as education and income) by CCDF subsidy receipt might be confounded by factors that simultaneously explain selection into CCDF and/or independently affect those outcomes. The mean difference in outcomes is subject to reverse causality and selection biases. Both the sign and magnitude of the mean differences that can be computed from data in Exhibit 6 should be taken cautiously as they might be affected by those biases.

[9] Notice that from a conceptual standpoint income can be considered both a predictor, and an outcome for subsidy receipt. Disentangling the variability in household income that causally explains CCDF participation, but is not explained by subsidy receipt (i.e., dealing with reverse causality), is a complex statistical task—one that has eluded subsidy researchers since the program was created. To what extent observed income today is induced by the expected benefit associated with CCDF subsidy use tomorrow is an unresolved question.

participants are heavily concentrated in the category of "Some college/no degree" (0.69) with a difference that is statistically significant from the group of CCDF non-participants (0.19). Similar results are obtained if we make those comparisons with reference to a CCDF variable measured up to 360 days before the survey interview date. Regarding the race and ethnicity of CCDF recipients, they appear to be more likely to be black or African-American, considerably less likely to be non-Hispanic white and with a similar proportion of Hispanics as households not receiving CCDF benefits. CCDF recipients are also more likely to have a child younger than sixty months old than non-recipients. The NSECE data files include some variables drawn from American Community Survey data to describe the communities in which survey households are located. Compared to non-recipients, CCDF participants tend to reside in communities with higher levels of poverty density.

Constructing the CCDF Receipt Variable: Does the Choice of Time before the NSECE Survey Matter?

Comparing the pairs of columns 1 and 2 with columns 3 and 4 by simple inspection in Exhibit 6, we can informally study whether the associations of the variables chosen with the CCDF measure differ with the time interval (time before the NSECE survey) for which subsidy receipt is measured (i.e., whether it matters if subsidy receipt is computed 30 or 360 days before the NSECE interview date).

The first pattern to notice is that the choice of time interval (30 or 360 days) affects the proportion characteristics of CCDF subsidy recipients, but has very little to no effect on the characteristics of non-participants. This result can occur because new households using CCDF are proportionally drawn from the multiple categories that define each dimension among CCDF non-recipients to become subsidy participants. The increased pool of CCDF participants is also small relative to the overall set of households that do not participate in the CCDF program. Consequently, when a small number of households start receiving CCDF subsidies between 30 and 360 days before the NSECE interview, they do not impact the overall proportions that characterize the attributes of non-participant households portrayed in the table.

Income to poverty ratios and educational attainment appear affected by the time interval for defining CCDF enrollment. We observe that, the broader the time interval that allows for subsidy receipt, the more households we have in the second and third brackets of income to poverty (i.e., in the 100% to <200%, and 200% to <300% poverty categories). Interestingly, a similar pattern of realignment across categories is observed for educational attainment of the survey respondent: as the time interval elongates, the fewer households we see in the category "Some college/no degree" and the more we see in the category defined as "Bachelor degree." Previous research has not acknowledged that the time dimension of how we define CCDF treatment affects the attributes of the group of treated observations. When using combined survey and administrative data, the overall magnitude of treatment effects of the CCDF program on both earnings and education of participants might differ according to what time frame the researcher chooses to define subsidy receipt.

Other variables in our analysis also show differences according to the time period used to define CCDF receipt, but less so. Only the racial composition and the number of households with more than one child ages 0-5 change. In the case of race and ethnicity we compared columns 1 to 3 to observe that only the proportion of white households changes importantly from 0.33 to 0.24, while the rest of the categories remain roughly the same. The proportion of

households that have at least one child ages 0-5 and use CCDF increases from 0.55 to 0.64 when CCDF receipt is defined based on 30 vs 360 days prior to the NSECE interview date.

The results in this section should all be taken as exploratory, as they are the result of a first attempt to link HH-NSECE and CCDF administrative records to draw broad methodological lessons. One striking finding of this exploration seems to be that some variables are likely affected by the choice of either 30 or 360 days for defining CCDF receipt, while others seem less sensitive to this choice. One hypothesis for why these patterns occur is a "dosage" effect; perhaps defining CCDF receipt over a longer time period includes more heterogeneous spells of CCDF exposure and so leads to heterogeneity in the outcomes we observe in the data.

CONCLUDING REMARKS

Since the CCDF program was initiated in 1996, one key challenge faced by researchers has been data limitations for accurately measuring the enrollment of eligible households in this program while also having adequate additional information to inform analyses. Combining survey data with administrative records can open new avenues for CCDF subsidy research. This report discusses new challenges and opportunities that arise as researchers combine survey data with CCDF administrative records to conduct subsidy research.

We highlight several methodological lessons for linking survey data with administrative data to study the CCDF program. The fact that researchers will almost always have partial (rather than full) consent of households for linkage gives rise to two challenging statistical problems: potential bias and likely undercount. The seemingly straightforward process of pairing the time of the NSECE Household interview with that of subsidy receipt in administrative records demands careful conceptualization of what the NSECE variables express when they are measured after and/or before CCDF subsidy receipt. As the present work pioneers an emerging approach for analysis of CCDF subsidy receipt, we have prioritized the methodological contributions over the substantive ones. (The latter are left for future work.)

We hope that the present report can serve as a methodological guide for future research that intends to develop and employ integrated data systems to overcome limitations that come from exclusively using either survey data or administrative data to conduct CCDF policy research.

References

Blau, David. "*Do Child Care Regulations Affect the Child Care and Labor Markets?*" Journal of Policy Analysis and Management 22 (3), Summer 2003: 443-465.

Blau, David and Janet Currie. "*Who's Minding the Kids?: Preschool, Day Care, and After School Care*," The Handbook of Education Economics, Finis Welch and Eric Hanushek (eds). New York: North Holland, 2006, v.2 1163-1278.

National Survey of Early Care and Education Project Team (2013). *National Survey of Early Care and Education: Summary Data Collection and Sampling Methodology*. OPRE Report #2013-46, Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Available at: http://www.acf.hhs.gov/programs/opre/research/project/national-survey-of-early-care-and-education-nsece-2010-2014

APPENDIX

Section A: The Illinois Childcare Assistance Program

The Illinois Childcare Assistance Program (CCAP) provides child care subsidies to Illinois' low-income working parents. Created in 1997 in the context of welfare reform, the program is partially funded by the Child Care Development Fund (CCDF).[10] In 2010, CCAP subsidized child care for approximately 130,000 children and is funded with federal and state resources amounting to $600 million.[11] Aligned with welfare reform priorities, CCAP aims to help low-income parents sustain employment. The program particularly targets low-income single mothers who are either dependent, or at risk of becoming dependent, on public assistance.

CCAP issues certificates or vouchers to parents upon request, primarily through resource and referral agencies that administer the program on behalf of the state of Illinois.[12] The program does not have a waiting list. Parents are entitled to child care for children from birth to age thirteen if they meet the program's eligibility criteria. (The criteria include an income threshold level. The level varies depending on family size.)[13] Parents must be employed when applying for the program, but some parents attending school and/or participating in job training activities are also eligible (priority is given to TANF recipients). Employed parents form, by far, the largest group of CCAP recipients. In FY 2002, 88 percent of CCAP participants entered the program under the employment eligibility category.

Eligibility for CCAP benefits is reevaluated every six months, and parents are required to notify the resource and referral agency handling the case of any change in work, family composition, and/or income status between redetermination periods.[14] Eligible households are required to share in the cost of care by making copayments, the size of which varies according to household income, size, the number of children in care, hours in care, and the type of provider chosen. CCAP subsidizes full time and part-time child care in licensed or license-exempt child care centers and homes, in-home nonrelative care, and care provided by relatives.

Section B: Missing Data on Income to Poverty

As explained in the main text, the household income to poverty ratio is a key variable in this analysis because it shows statistical significance as a predictor of CCDF subsidy utilization. Since approximately 3.5 percent of its values are missing (non-response) we studied whether data on this variable were missing in a way that is uncorrelated with other key variables: in the survey, namely respondent's education, race, as well as presence of at least one non-parent adult and of at least one child younger than five in the household.

---

[10] A complete description of the history and evolution of the CCDF program can be found in the 2008 edition of Background Material and Data on the Programs within the Jurisdiction of the Committee on Ways and Means (Green Book). Blau (2003) and Blau and Currie (2006) provide excellent detailed analysis of the CCDF program. Our characterization of the program is based on those sources. The details of the Illinois CCAP presented here are compiled from reviews of the yearly "Plan for CCDBG services in Illinois" from 1998 to 2011.

[11] According to "Plan for CCDBG services in Illinois, for the period 10/1/09–9/30/11," funding for CCAP included the following sources: Child Care Development Fund ($207,581,165); direct Federal TANF spending on child care ($168,034,659); state CCDBG maintenance of effort funds ($56,873,825); state matching funds ($71,736,826); extra general revenue ($249,239,418).

[12] The CCAP also works with selected licensed child care providers who help parents access CCAP.

[13] In 1997 the income threshold for a family of three was 166% of the federal poverty line (FPL) while in 2010 the threshold was 200% of the FPL.

[14] Workers paid in cash and the self-employed are reassessed for eligibility every three months.

Exhibit A1 shows the regression outcome from a binary outcome model where the dependent variable is an indicator for whether the household data on income to poverty ratio were missing (the binary indicator variable turns one if a household has missing income to poverty data, and zero otherwise). The table evidences that there is no statistically significant relationship (neither individually nor jointly) between having missing data on income to poverty ratio and the other variables in the analysis.

Exhibit A1. Studying Missing Values in Income to Poverty Ratio Variable

Variable	Coefficient (SE below)
HHs with any child aged 0 to < 60 months	0.0178
	(0.0133)
One non parent adult in HH	0.0020
	(0.0139)
8th grade or less	0.0311
	(0.1814)
9th- 12th grade	0.0271
	(0.1805)
HS or GED	0.0667
	(0.1802)
Some college/no degree	0.0473
	(0.1805)
Associate degree	0.0143
	(0.1813)
Bachelor degree	0.0441
	(0.1810)
Graduate degree	0.0579
	(0.1822)
White	-0.0153
	(0.0214)
Black	0.0029
	(0.0225)
Hispanics	-0.0016
	(0.0220)
Constant	-0.0182
	(0.1817)
Observations	755
R^2	0.0127
Log likelihood	232.6797

Section C: Inverse Probability Weighting

We implemented an inverse probability weights (IPW) approach to account for the truncated nature of the sample of respondents who consented to be matched to administrative records in Illinois. The intuition behind the strategy is simple and consists in weighting more heavily those households that, having given consent, are more alike to those households that did not give

consent. Re-weighting the data in this way ensures that the population counts (both households and children) are representative of the Illinois population.

The IPW scheme was implemented in two stages, the first of which was to determine the average probability of consent (PC) for all households in the sample to calculate the ratio 1/(1-PC)=IPW. This IPW ratio represents the household-level inverse of the probability of not giving consent. The last stage in the implementation of the IPW scheme was to make the IPW weights for each household enter the sampling design by multiplying the NSECE household sampling weights.[15]

In Exhibit A2, we show the implications derived from reweighting the data using the IPW strategy described to recover population counts. The table tabulates descriptive statistics as well as sample and population counts of the sample of Illinois households that consented across four categories of the household income to FPL ratio. As expected, the population counts are smaller when we use the original NSECE HH survey weights than when we use weights that account for the truncated nature of the consent sample (counting approx. 900,000, instead of 1,200,000 households). Notice also, that the mean and standard deviations are not affected by the choice of weights. Undercounting households that would have resulted from solely using NSECE weights is addressed by the IPW strategy.

Exhibit A2. Descriptive Statistics of the Consent Sample: Income to Federal Poverty Level (FPL) Ratio with and without IPW Weights

	HH Weights				HH+IPW Weights			
Income to FPL Ratio	**Mean**	**SE**	**N Sample**	**N Pop.**	**Mean**	**SE**	**N Sample**	**N Pop.**
<100% poverty	0.26	0.026	570	898,000	0.26	0.026	570	1,190,000
>=100 to <200% poverty	0.24	0.028	570	898,000	0.24	0.028	570	1,190,000
>=200 to 300% poverty	0.17	0.028	570	898,000	0.17	0.028	570	1,190,000
>=300% poverty	0.31	0.034	570	898,000	0.31	0.034	570	1,190,000

[15] Notice that, in this analysis, IPW is constant across all households since there are not statistically significant differences in observables factors characterizing individual households in explaining the consent decision. The effect of re-weighting on this analysis is simply to adjust the total count of households to cover the full target population.